**

**IMPLEMENTASI METODE SELEKSI FITUR PADA DATA TIME SERIES MULTIVARIAT UNTUK PREDIKSI CURAH HUJAN**

RIZKA ANNISA KURNIA SARI

NRP 051115 40000 114

Dosen Pembimbing

SARWOSRI, S.Kom., M.T.

AHMAD SAIKHU, S.Si., M.T.

DEPARTEMEN INFORMATIKA

Fakultas Teknologi Informasi dan Komunikasi

Institut Teknologi Sepuluh Nopember

Surabaya 2019

TUGAS AKHIR – K141502

*[Halaman ini sengaja dikosongkan]*



**IMPLEMENTASI METODE SELEKSI FITUR PADA DATA TIME SERIES MULTIVARIAT UNTUK PREDIKSI CURAH HUJAN**

RIZKA ANNISA KURNIA SARI

NRP 051115 40000 114

Dosen Pembimbing

SARWOSRI, S.Kom., M.T.

AHMAD SAIKHU, S.Si., M.T.

DEPARTEMEN INFORMATIKA

Fakultas Teknologi Informasi dan Komunikasi

Institut Teknologi Sepuluh Nopember

Surabaya 2019

TUGAS AKHIR – K141502

*[Halaman ini sengaja dikosongkan]*



**IMPLEMENTATION OF SELECTION METHODS IN MULTIVARIATE TIME SERIES DATA FOR PREDICTING RAINFALL**

FINAL PROJECT – K141502

RIZKA ANNISA KURNIA SARI

NRP 051115 40000 114

Advisor

SARWOSRI, S.Kom., M.T.

AHMAD SAIKHU, S.Si., M.T.

INFORMATICS DEPARTMENT

Faculty of Information and Communication Technology

Institut Teknologi Sepuluh Nopember

Surabaya 2019

*[Halaman ini sengaja dikosongkan]*

# LEMBAR PENGESAHAN

**IMPLEMENTASI METODE SELEKSI FITUR PADA DATA TIME SERIES MULTIVARIAT UNTUK PREDIKSI CURAH HUJAN**

**TUGAS AKHIR**

Diajukan Guna Memenuhi Salah Satu Syarat   
Memperoleh Gelar Sarjana Komputer  
pada   
Bidang Studi Dasar dan Terapan Komputasi  
Program Studi S-1 Departemen Informatika  
Fakultas Teknologi Informasi dan Komunikasi  
Institut Teknologi Sepuluh Nopember

Oleh:  
**RIZKA ANNISA KURNIA SARI**NRP: 051115 40000 114

Disetujui oleh Dosen Pembimbing Tugas Akhir:

|  |  |
| --- | --- |
| SARWOSRI, S.Kom., M.T. NIP: 19760809 200112 2 001 | ................................ (pembimbing 1) |
|  |  |
| AHMAD SAIKHU, S.Si., M.T. NIP: 19710718 200604 1 001 | ................................ (pembimbing 2) |
|  |  |

**Surabaya  
JANUARI 2019**

*[Halaman ini sengaja dikosongkan]*

**IMPLEMENTASI METODE SELEKSI FITUR PADA DATA TIME SERIES MULTIVARIAT UNTUK PREDIKSI CURAH HUJAN**

Nama Mahasiswa : Rizka Annisa Kurnia Sari

NRP : 051115 40000 114

Departemen : Informatika FTIK-ITS

Dosen Pembimbing 1 : Sarwosri, S.Kom., M.T.

Dosen Pembimbing 2 : Ahmad Saikhu, S.Si., M.T.

# ABSTRAK

*Gejala alam yang terjadi di dunia ini seperti cuaca, iklim dan musim merupakan salah satu faktor yang sangat berpengaruh kepada keberlangsungan kehidupan manusia.* *Selain memberikan berbagai anugrah dan kebermanfaatan bagi manusia, gejala alam yang dapat terjadi juga bisa mendatangkan bahaya atau bahkan bisa jadi menyebabkan terjadinya bencana. Oleh karena itu data/informasi yang didapatkan saat ini dapat dimanfaatkan untuk memprediksi kemungkinan yang terjadi dan mengurangi terjadinya resiko akibat bencana alam dan melakukan tindakan mitigasi. Berdasarkan studi, curah hujan di Indonesia dipengaruhi oleh berbagai faktor yaitu lokal (cuaca) dan iklim regional (global). Oleh karenanya penentuan prediktor yang relevan adalah penting untuk efisiensi model penduga curah hujan. Meskipun beberapa penelitian pemodelan iklim di suatu wilayah/negara memiliki akurasi tinggi, belum tentu model tersebut sesuai untuk wilayah lainnya dikarenakan berbagai faktor.Dalam penelitian ini melakukan pendekatan yang melibatkan sejumlah fitur cuaca dan iklim dalam ruang lingkup lokal dan global dengan memperhitungkan faktor time-lag (temporal) dan lokasi (spasial). Dengan demikian, prediksi curah hujan baik dari segi prediktor yang simultan maupun model Long Short-Term Memories, prediksi yang akan dihasilkan akan lebih akurat sehingga dapat digunakan sebagai informasi perencanaan di berbagai bidang. Pemilihan prediktor dengan hanya melibatkan fitur relevan (signifikan) menjadikan proses komputasi lebih efisien sehingga informasi yang didapat untuk pengambilan keputusan akan didapat lebih cepat.*

***Kata kunci: cuaca, iklim, spatio temporal, time series, detrended partial cross correlation, neural network, long short term memories.***

**IMPLEMENTASI METODE SELEKSI FITUR PADA DATA TIME SERIES MULTIVARIAT UNTUK PREDIKSI CURAH HUJAN**

Nama Mahasiswa : Rizka Annisa Kurnia Sari

NRP : 051115 40000 114

Departemen : Informatika FTIK-ITS

Dosen Pembimbing 1 : Sarwosri, S.Kom., M.T.

Dosen Pembimbing 2 : Ahmad Saikhu, S.Si., M.T.

# ABSTRACT

*Natural phenomena that occur in the world such as weather, climate and seasons are one of the factors that greatly influence the sustainability of human life. In addition to giving various gifts and benefits to humans, natural phenomena that can occur can also bring harm or may even cause disasters. Therefore data / information obtained at this time can be used to predict the possibilities that occur and reduce the occurrence of risks due to natural disasters and take mitigation actions. Based on the study, rainfall in Indonesia is influenced by various factors, namely local (weather) and regional (global) climate. Therefore the determination of relevant predictors is important for the efficiency of the rainfall estimator model. Although some climate modeling studies in a region / country have high accuracy, it is not necessarily the model that is suitable for other regions due to various factors. In this study an approach that involves a number of weather and climate features in local and global scope taking into account time-lag factors (temporal) and location (spatial). Thus, predictions of rainfall both in terms of simultaneous predictors and Long Short-Term Memories models, predictions to be produced will be more accurate so that they can be used as planning information in various fields. The selection of predictors by only involving relevant features (significant) makes the computing process more efficient so that the information obtained for decision making will be obtained more quickly.*

*.*

***Keywords: weather, climate, spatio temporal, time series, detrended partial cross correlation, neural network, long short term memories.***

# KATA PENGANTAR



Puji syukur penulis panjatkan ke hadirat Allah SWT karenaa atas karunia dan rahmat-Nya penulis dapat menyelesaikan tugas akhir yang berjudul:

**IMPLEMENTASI METODE SELEKSI FITUR PADA DATA TIME SERIES MULTIVARIAT UNTUK PREDIKSI CURAH HUJAN**

Melalui lembar ini, penulis ingin menyampaikan ucapan terima kasih dan penghormatan yang sebesar-besarnya kepada:

1. Kedua orang tua penulis, Mama dan Papa yang selalu menemani, mengingatkan, mendoakan, dan mejadi #1 *Support System* bagi penulis.
2. Mbak Fika, Mas Rheza, Mas Bayu, Ara dan keluarga besar yang selalu memberikan doa serta berbagai macam bentuk dukungan baik berupa materil dan moril kepada penulis untuk menyelesaikan tugas akhir ini.
3. Ibu Sarwosri, S.Kom., M.T. selaku dosen pembimbing tugas akhir pertama yang telah memberikan masukkan serta koreksi dalam pengerjaan tugas akhir. Bapak Ahmad Saikhu, S.Si., M.T. selaku dosen pembimbing tugas akhir kedua yang telah membimbing dan sangat banyak membantu pengerjaan program tugas akhir ini.
4. Bapak dan Ibu dosen, Karyawan serta seluruh civitas Departemen Informatika yang telah memberikan pelajaran dan pengalaman selama menjadi mahasiswa di Departemen Informatika.
5. Cumi Cumi Keju (Neny, Devi, Sarah, Nada) yang telah menjadi teman penghilang *stress* dan penat yang selalu ada baik sebelum hingga pengerjaan tugas akhir ini selesai.
6. Zakiya, Nindy, Bonbon, Irsa, Ufa, Cynde dan kawan-kawan TC15, C1F, serta seluruh anggota HMTC yang sudah menemani, mendukung dan memberikan pelajaran dalam organisasi maupun perkuliahan.
7. Teman-teman se-*rumah* di lab. LP. Bonbon, Raca, Brian. Yuuta, Irsyad, Mala, Ghisa, Chael, John, Mas Afiif, Mas Hari, Mas Nafiar dan yang lainnya untuk semangat, dukungan serta rasa kekeluargaan yang ada.
8. Serta pihak-pihak lain yang namanya tidak dapat penulis sebutkan satu per satu.

Bagaimanapun juga penulis telah berusaha sebaik-baiknya dalam menyelesaikan tugas akhir ini. Namun, penulis mohon maaf apabila terdapat kekurangan ataupun kesalahkan yang penulis lakukan. Kritik dan saran yang membangun dapat disampaikan sebagai bahan perbaikan untuk ke depannya.

Surabaya, Januari 2019

Rizka Annisa Kurnia Sari

# DAFTAR ISI

[LEMBAR PENGESAHAN vii](#_Toc534191276)

[ABSTRAK ix](#_Toc534191277)

[ABSTRACT xi](#_Toc534191278)

[KATA PENGANTAR xiii](#_Toc534191279)

[DAFTAR ISI xv](#_Toc534191280)

[DAFTAR GAMBAR xix](#_Toc534191281)

[DAFTAR TABEL xxi](#_Toc534191282)

[DAFTAR KODE SUMBER xxiii](#_Toc534191283)

[BAB I PENDAHULUAN 1](#_Toc534191284)

[1.1 Latar Belakang 1](#_Toc534191285)

[1.2 Rumusan Permasalahan 2](#_Toc534191286)

[1.3 Batasan Permasalahan 2](#_Toc534191287)

[1.4 Tujuan 3](#_Toc534191288)

[1.5 Manfaat 4](#_Toc534191289)

[1.6 Metodologi 4](#_Toc534191290)

[1.7 Sistematika Penulisan 5](#_Toc534191291)

[BAB II DASAR TEORI 7](#_Toc534191292)

[2.1 Cuaca dan Iklim 7](#_Toc534191293)

[2.2 *El Nino-Southern Oscillation* (ENSO) 7](#_Toc534191294)

[*2.3* *Bahasa R* 7](#_Toc534191295)

[*2.4* *Keras* 8](#_Toc534191296)

[2.5 *Multivariate Time Series* (MTS) 8](#_Toc534191297)

[2.6 *Detrended Parcial Cross-Correlation Analysis* (DPCCA) 9](#_Toc534191298)

[2.7 *Long Short Term Memories* (LSTM) 11](#_Toc534191299)

[2.8 RMSprop 12](#_Toc534191300)

[*2.9* *Categorical Cross-Entropy* 13](#_Toc534191301)

[BAB III ANALISIS DAN PERANCANGAN SISTEM 15](#_Toc534191302)

[3.1 Analisis Metode Secara Umum 15](#_Toc534191303)

[3.2 Perancangan Data 16](#_Toc534191304)

[3.2.1. Spatio Temporal 17](#_Toc534191305)

[3.2.2. Data *Training* 17](#_Toc534191306)

[3.2.3. Data *Testing* 18](#_Toc534191307)

[3.3 Perancangan Proses 18](#_Toc534191308)

[3.3.1. Metode DPCCA 18](#_Toc534191309)

[3.3.2. Metode LSTM 20](#_Toc534191310)

[BAB IV IMPLEMENTASI 21](#_Toc534191311)

[4.1 Lingkungan Implementasi 21](#_Toc534191312)

[4.2 Implementasi Proses 21](#_Toc534191313)

[4.2.1. Implementasi Metode *Moving Average* 22](#_Toc534191314)

[4.2.2. Implementasi Metode *Support Vector Regression* 23](#_Toc534191315)

[4.2.3. Implementasi Metode *Support Vector Regression – Genetic Algorithm* 26](#_Toc534191316)

[4.2.4. Implementasi Metode *Support Vector Regression – Particle Swarm Optimization* 31](#_Toc534191317)

[BAB V PENGUJIAN DAN EVALUASI 33](#_Toc534191318)

[5.1 Lingkungan Pengujian 33](#_Toc534191319)

[5.2 Data Uji Coba 33](#_Toc534191320)

[5.3 Skenario Uji Coba 33](#_Toc534191321)

[5.4 Skenario Pengujian 1 36](#_Toc534191322)

[5.4.1. Skenario Pengujian 1.1 37](#_Toc534191323)

[5.4.2. Skenario Pengujian 1.2 38](#_Toc534191324)

[5.4.3. Skenario Pengujian 1.3 39](#_Toc534191325)

[5.4.4. Skenario Pengujian 1.4 40](#_Toc534191326)

[5.5 Skenario Pengujian 2 41](#_Toc534191327)

[5.5.1. Skenario Pengujian 2.1 41](#_Toc534191328)

[BAB VI KESIMPULAN DAN SARAN 44](#_Toc534191329)

[6.1. Kesimpulan 44](#_Toc534191330)

[6.2. Saran 45](#_Toc534191331)

[DAFTAR PUSTAKA 46](#_Toc534191332)

[LAMPIRAN 48](#_Toc534191333)

[BIODATA PENULIS 56](#_Toc534191334)

*[Halaman ini sengaja dikosongkan]*

# DAFTAR GAMBAR

[Gambar 2. 1 Contoh data time series yang diperoleh dari Climexp pada tahun 1987-2016 9](#_Toc534190548)

[Gambar 2. 2 Dendogram CCF antara curah hujan dan temperatur 10](#_Toc534190549)

[Gambar 2. 3 Modul berulang dalam LSTM 11](#_Toc534190550)

[Gambar 3. 1 Diagram alir implementasi metode secara umum 16](#_Toc534190574)

[Gambar 3. 2 Stacked layer LSTM 20](#_Toc534190575)

*[Halaman ini sengaja dikosongkan]*

# DAFTAR TABEL

[Tabel 3. 1 Jarak antar stasiun 17](#_Toc534190627)

[Tabel 3. 2 Detil jumlah baris dan kolom data training 18](#_Toc534190628)

[Tabel 3. 3 Detil jumlah baris dan kolom data testing 18](#_Toc534190629)

*[Halaman ini sengaja dikosongkan]*

# DAFTAR KODE SUMBER

*[Halaman ini sengaja dikosongkan]*

# BAB I PENDAHULUAN

## Latar Belakang

Gejala alam yang terjadi di dunia ini seperti cuaca, iklim dan musim merupakan salah satu faktor yang sangat berpengaruh kepada keberlangsungan kehidupan manusia. Selain memberikan berbagai anugrah dan kebermanfaatan bagi petani, nelayan, pelayar, dsb. Gejala alam tersebut juga bisa mendatangkan bahaya atau bahkan bisa jadi menyebabkan terjadinya bencana. diantaranya wabah, penyakit, banjir, angin kencang, tanah longsor, kekeringan, kebakaran, dsb. Oleh karena itu data/informasi yang didapatkan saat ini dapat dimanfaatkan untuk memprediksi kemungkinan yang terjadi dan mengurangi terjadinya resiko akibat bencana alam dan melakukan tindakan mitigasi.

Berdasarkan studi, curah hujan di Indonesia dipengaruhi oleh berbagai faktor yaitu lokal (cuaca) dan iklim regional (global). Oleh karenanya penentuan prediktor yang relevan adalah penting untuk efisiensi model penduga curah hujan. Meskipun beberapa penelitian pemodelan iklim di suatu wilayah/negara memiliki akurasi tinggi, belum tentu model tersebut sesuai untuk wilayah lainnya dikarenakan berbagai faktor. Hal ini menyebabkan bahwa pemodelan untuk iklim dan prediksi hujan menjadi khas untuk suatu wilayah. Model iklim dan curah hujan wilayah tropis yang mampu mensimulasikan dan memprediksi curah hujan di Indonesia secara efektif belum ditemukan.

Salah satu pendekatan model empiris adalah metode ARIMA yang memiliki kelemahan kurang sesuai untuk data non-linier sehingga akurasinya rendah. Untuk mengatasi hal ini, salah satu pilihan adalah menggunakan model ANN yang juga memiliki kelemahan bahwa untuk pola linier, tidak menjamin akurasi yang optimal dan bergantung pada karakteristik data.

Oleh karenanya dalam penelitian ini diusulkan sebuah pendekatan baru model penduga curah hujan wilayah Zona Musim di Indonesia melalui kombinasi proses ekstrasi dan seleksi variabel prediktor serta penggunaan *Long Short Term Memory* (LSTM). Pendekatan yang dimaksud yaitu dengan melibatkan sejumlah fitur cuaca dan iklim dalam ruang lingkup lokal dan global dengan memperhitungkan faktor *time-lag* (temporal) dan lokasi (spasial). Dengan demikian, prediksi curah hujan baik dari segi prediktor yang simultan maupun model LSTM, prediksi yang akan dihasilkan akan lebih akurat sehingga dapat digunakan sebagai informasi perencanaan di berbagai bidang. Pemilihan prediktor dengan hanya melibatkan fitur relevan (signifikan) menjadikan proses komputasi lebih efisien sehingga informasi yang didapat untuk pengambilan keputusan akan didapat lebih cepat.

## Rumusan Permasalahan

Rumusan masalah yang diangkat dalam tugas akhir ini adalah sebagai berikut:

1. Bagaimana melakukan ekstraksi fitur time series multivariat dari variabel prediktor yang berhubungan dengan variabel curah hujan?
2. Bagaimana melakukan seleksi prediktor lokal dan global untuk model penduga curah hujan sehingga terpilih fitur yang relevan secara linier dan non linier?
3. Bagaimana membangun model penduga curah hujan berbasis LSTM yang mampu mengintegrasikan variabel lokal/global, linier/non-linier, dan spatio-temporal?

## Batasan Permasalahan

Batasan masalah pada tugas akhir ini antara lain:

1. *Tools* yang digunakan adalah RStudio-64bit.
2. Data yang digunakan adalah data curah hujan dari BMKG (http://dataonline.bmkg.go.id/) untuk wilayah Surabaya, meliputi Stasiun Meteorologi Juanda, Stasiun Meteorologi Perak I dan Stasiun Meteorologi Maritim Perak II.
3. Periode data observasi yang digunakan mulai dari 01 Januari 1987 – 31 Desember 2016 (30 tahun).
4. Jumlah variable yang digunakan sebanyak 12 variabel, yang terdiri dari 8 variabel prediktor lokal dan 4 variabel prediktor global.
5. Variabel prediktor lokal yang digunakan dari data BMKG adalah suhu minimum, suhu maksimum, suhu rata-rata, kelembaban rata-rata, curah hujan, lama penyinaran, kecepatan angin rata-rata dan kecepatan angina terbesar.
6. Variabel prediktor global yang digunakan dari data KNMI (http://climexp.knmi.nl/) adalah NINO12, NINO 3, NINO3.4 dan NINO4)
7. Metode yang digunakan untuk melakukan seleksi fitur adalah *Detrended Partial Cross-Correlation Analysis* (DPCCA) dan semi-DPCCA.
8. Metode yang digunakan untuk model penduga curah hujan adalah LSTM.
9. Hasil prediksi yang ditemukan adalah prediksi pada data harian di tahun berikutnya.

## Tujuan

Tujuan dari pembuatan tugas akhir ini adalah membangun model penduga curah hujan dengan sebuah pendekatan baru model penduga curah hujan wilayah Zona Musim di Indonesia melalui kombinasi proses ekstrasi dan seleksi variabel prediktor serta penggunaan *Long Short Term Memory* (LSTM) untuk meningkatkan akurasi prediksi.

Proses ekstraksi dan seleksi fitur dimaksud adalah menemukan kandidat fitur *time-lags* dari prediktor lokal/global menggunakan metode CCF, dan melakukan seleksi prediktor baik hubungan linier maupun non-linier menggunakan metode FCBF. Hal ini bertujuan untuk memperoleh input prediktor yang relevan.

## Manfaat

Manfaat dari pembuatan tugas akhir ini adalah:

1. Tersedianya metode seleksi prediktor yang mampu mengidentifikasi relevansi linier dan non-linier terhadap variabel respon.
2. Tersedianya model penduga curah hujan dengan periode harian yang melibatkan sejumlah variabel input secara simultan, baik variabel lokal/global, waktu maupun wilayah.
3. Dapat digunakan sebagai generalisasi untuk wilayah ZoM di luar Jawa Timur untuk memprediksi curah hujan harian di mana 84% wilayah Indonesia adalah ZoM.

## Metodologi

Langkah-langkah yang ditempuh dalam pengerjaan tugas akhir ini adalah sebagai berikut:

1. **Studi literatur**

Pada studi literatur ini akan dipelajari sejumlah referensi yang relevan terhadap tugas akhir yang akan dikerjakan. Studi literatur ini didapatkan dari buku, internet serta materi-materi kuliah yang berhubungan dengan metode yang akan digunakan. Hal-hal yang akan dipelajari yaitu mengenai R, *Long Short Term Memory* (LSTM) dan *Detrended Partial Cross-Correlation Analysis* (DPCCA).

1. **Analisis dan desain metode**

Metode yang akan dikembangkan merupakan penggabungan metode linier dan non-linier dari DPCCA dengan penggunaan spatio-temporal. Desain metode ini dibutuhkan untuk mengetahui predictor yang berpengaruh untuk dilakukan *forecast* dengan metode LSTM.

1. **Implementasi**

Pengembangan dari metode yang akan dibuat pada tugas akhir ini akan menggunakan tools berupa RStudio 64bit karena telah menyediakan *library* yang dibutuhkan untuk menjalankan metode DPCCA dan LSTM.

1. **Pengujian dan evaluasi**

Pada tahap ini dilakukan pengujiian dari model yang telah dibuat. Langkah pertama adalah melakukan pengujian menggunakan dataset masing-masing stasiun yang sudah diseleksi fitur secara linier. Langkah kedua, melakukan pengujian dataset masing-masing stasiun yang sudah diseleksi fitur secara non-linier. Selanjutnya melakukan penggabungan 3 lokasi. Hasil akhir dari pengujian tersebut akan dievaluasi menggunakan *Mean Absolute Error* (MAE) dan *Categorical Cross-Entropy* untuk melihat model mana yang memberikan hasil dengan tingkat akurasi tertinggi.

1. **Penyusunan buku Tugas Akhir**

Pada tahap ini dilakukan proses dokumentasi dan pembuatan laporan dari seluruh konsep, tinjauan pustaka, metode, implementasi, proses yang telah dilakukan, pengujian, evaluasi dan hasil-hasil yang telah didapatkan selama pengerjaan tugas akhir.

## Sistematika Penulisan

Buku tugas akhir ini bertujuan untuk mendapatkan gambaran dari pengerjaan tugas akhir. Selain itu, diharapkan dapat berguna untuk pembaca yang tertarik untuk melakukan pengembangan lebih lanjut. Secara garis besar, buku tugas akhir terdiri atas beberapa bagian seperti berikut ini:

**Bab I Pendahuluan**

Bab ini berisi latar belakang masalah, tujuan dan manfaat pembuatan tugas akhir, permasalahan, batasan masalah, metodologi yang digunakan, dan sistematika penyusunan tugas akhir.

**Bab II Dasar Teori**

Bab ini menjelaskan beberapa teori yang dijadikan penunjang dan berhubungan dengan pokok pembahasan yang mendasari pembuatan tugas akhir.

**Bab III Analisis dan Perancangan Sistem**

Bab ini membahas mengenai perancangan sistem yang akan dibangun. Perancangan sistem meliputi perancangan data dan alur proses dari sistem itu sendiri.

**Bab IV Implementasi**

Bab ini berisi implementasi dari perancangan sistem yang telah ditentukan sebelumnya.

**Bab V Pengujian dan Evaluasi**

Bab ini membahas pengujian dari metode yang ditawarkan dalam tugas akhir untuk mengetahui kesesuaian metode dengan data yang ada.

**Bab VI Kesimpulan dan Saran**

Bab ini berisi kesimpulan dari hasil pengujian yang telah dilakukan. Bab ini juga membahas saran-saran untuk pengembangan sistem lebih lanjut.

**Daftar Pustaka**

Merupakan daftar referensi yang digunakan untuk mengembangkan tugas akhir.

**Lampiran**

Merupakan bab tambahan yang berisi data atau daftar istilah yang penting pada tugas akhir ini.

# BAB II DASAR TEORI

Bab ini membahas teori-teori yang menjadi dasar pembuatan tugas akhir.

## Cuaca dan Iklim

Cuaca adalah keadaan udara (atmosfir bumi) pada saat tertentu di wilayah tertentu yang relatif sempit dan pada jangka waktu yang singkat. Cuaca terjadi karena suhu dan kelembaban yang berbeda antara satu tempat dengan tempat lainnya.[6]

Iklim adalah keadaan cuaca rata-rata dalam waktu yang relatif lama dan meliputi wilayah yang lebih luas. Iklim di suatu tempat di bumi dipengaruhi oleh letak geografis dan topografi tempat tersebut. Pengaruh posisi relatif matahari terhadap suatu tempat di bumi menimbulkan musim, suatu penciri yang membedakan iklim satu dari yang lain.

## *El Nino-Southern Oscillation* (ENSO)

Hujan di Indonesia dipengaruhi oleh iklim global, diantaranya *El Nino-Southern Oscillation* (ENSO) di mana besarnya pengaruh beragam. Pengaruh itu sangat besar pada daerah yang memiliki pola hujan monsoon. El Nino merupakan salah satu bentuk penyimpangan iklim di Samudra pasifik yang ditandai dengan kenaikan *Sea Surface Temperature* (SST) di daerah katulistiwa bagian tengah dan timur. El Nino terdiri dari Nino 12, Nino 3, Nino 3.4 dan Nino 4.

## *Bahasa R*

R adalah bahasa pemrograman yang menyajikan *environment* untuk komputasi statistik dan grafik. R menyediakan berbagai macam statistik (pemodelan linier dan non-linier, uji statistik klasik, analisis deret waktu, klasifikasi, pengelompokan, dsb) dan teknik grafis, dan masih banyak fitur yang dapat diperluas. Kelebihan yang ada pada bahasa R adalah hasil plot dengan kualitas publikasi baik, serta kemudahan penggunaan simbol dan formula matematika.

## *Keras*

Keras adalah API *neural network* tingkat tinggi yang dikembangkan dengan fokus untuk mengaktifkan eksperimen secara cepat. Keras memiliki fitur utama berikut:

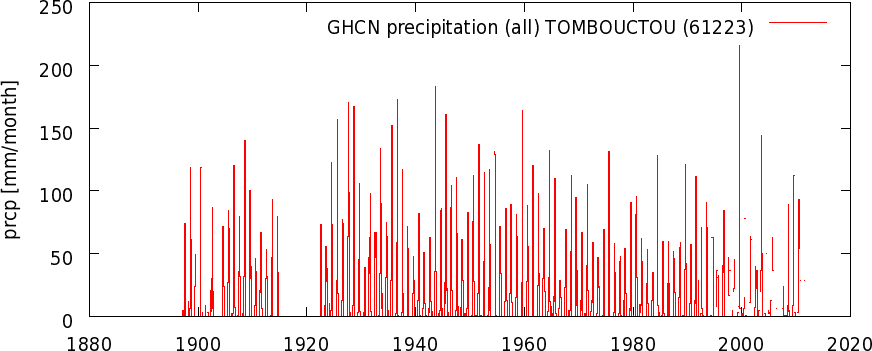
1. Mengizinkan kode yang sama dijalankan di CPU atau GPU tanpa cacat.
2. API yang *user-friendly* untuk memudahkan pembatan prototipe model *deep learning.*
3. Men*-support built-in* yang bisa digunakan untuk *convolutional networks* (untuk computer visi), *recurrent networks* (untuk pemrosesan sekuen) dan bisa untuk kombinasi dari keduanya.
4. Men-*support* arbitrary network architectures: *multi-input* atau *multi-output models*, *layer sharing*, *model sharing,* dsb. Dalam hal ini, keras dapat digunakan untuk membuat model *deep learning*, mulai dari *memory network* hingga *neural turing machine*.
5. Mampu berjalan di atas berbagai *back-end* termasuk TensorFlow, CNTK, atau Theano.

## *Multivariate Time Series* (MTS)

*Time Series* adalah representasi data yang terurut berdasarkan waktu kejadian dengan interval waktu yang sama di mana antar nilai saling berhubungan[2]. Data time series dapat dinotasikan sebagai berikut:

Xi(t); [I = 1, 2, 3, …n, t = 1, 2, 3, ….., m]

Jika nilai n > 2, disebut sebagai MTS, sementara apabila n = 1 dinamakan *Uniavriate Time Series* (UTS). Pada gambar 8.1 merupakan contoh data *time series*.

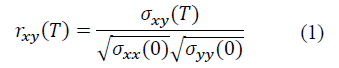


Gambar 2. 1 Contoh data time series yang diperoleh dari Climexp pada tahun 1987-2016

MTS adalah prediksi nilai kelas target berdasarkan sejumlah variabel prediktor di mana data juga terurut berdasarkan waktu kejadian dengan interval waktu yang sama[3]. Antar variabel prediktor terdapat hubungan, baik dalam periode pengamatan ke-t maupun ke (t-k). Dalam analisis time series, elemen penting adalah pola hubungan yang bersifat *time-delay* di antara variabel prediktor maupun antara variabel prediktor dengan label kelas.

## *Detrended Parcial Cross-Correlation Analysis* (DPCCA)

DPCCA merupakan Untuk mengidentifikasi hubungan antar variabel berdasarkan *time-lags* tertentu, alat analisis yang umum digunakan adalah Cross Correlation Function (CCF). Untuk menghitung CCF digunakan persamaan (1).



di mana rxy(T) adalah korelasi variabel x dan variabel y pada periode ke T dan σxy adalah kovariansi(x,y), σxx adalah variansi(x) dan σyy adalah variansi(y).

Dalam hubungan antara dua time series (yt dan xt), series yt mungkin berhubungan dengan *lags* yang lalu dari series-x. Sampel CCF membantu mengidentifikasi *lags* dari variabel x yang mungkin merupakan kandidat prediktor dari yt. Sampel CCF didefinikan sebagai himpunan korelasi sampel antara xt+h dan yt untuk h = 0, ±1, ±2, ±3, dan seterusnya. Nilai *negative* dari h menunjukkan bahwa korelasi antara variabel x pada waktu sebelum t dan variabel y pada saat t. Misalnya, h = −2 berarti bahwa nilai korelasi yang dihitung adalah xt-2 and yt. Hasil dari nilai korelasi dapat diuji dengan membandingkan nilai korelasi silang dengan *standard error* 1⁄√𝑛 − 𝑘. Pada gambar 8.2 adalah contoh dendogram dari CCF. Sedangkan 2 garis horizontal menunjukkan batas signifikansi. ­[4]­[5]



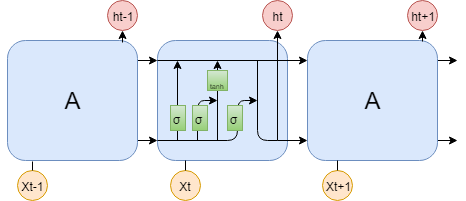
Gambar 2. 2 Dendogram CCF antara curah hujan dan temperatur

Karena CCF memiliki keterbatasan hanya untuk identifikasi pengaruh tenggang waktu yang bersifat linier dan stasioner, cenderung redundan dan bersifat *greedy* [12], maka metode lain dikembangkan, yaitu DCCA/DPCCA. Keunggulan metode ini adalah kemampuan identifikasi hubungan diantara dua variabel yang bersifat linier/non-linier dan non-stasioner sehingga dapat digunakan untuk memperoleh kandidat variabel prediktor lebih lengkap.

## *Long Short Term Memories* (LSTM)

*Long Short Term Memories* (LSTM) adalah unit bangunan untuk lapisan *Recurrent Neural Network* (RNN). LSTM sendiri di desain untuk mengatasi permasalahan *error* pada proses *back-flow*. Jika dibandingkan dengan RNN sederhana dengan *cell* yang berisi 1 *layer neuron* dengan fungsi aktivasi tanh. Dengan itu, dapat disimpulkan bahwa RNN biasa tidak cocok untuk mempelajari pola data berurut yang panjang.

Pada LSTM, isi *cell* menjadi lebih kompleks dari yang sekedar 1 *layer neuron* dan inilah yang menjadikan LSTM bisa mempelajari pola panjang dari data berurut karena situasi *vanishing gradient* dicegah. Yang membedakan LSTM dengan RNN hanya isi cellnya yang berbeda, selebihnya se!ara prinsip LSTM dengan RNN sama.



Gambar 2. 3 Modul berulang dalam LSTM

Pada gambar 2. 3 dijelaskan bahwa persegi panjang kuning melambangkan *layer neuron* (jadi bisa ada 2 atau lebih *neuron* pada masing-masing persegi panjang kuning tersebut jika *hidden neuron* yang dideklarasikan adalah 2 atau lebih) sedangkan bulatan berwarna merah muda melambangkan operasi *element-wise*. Panah hitam berkelok-kelok ini melambangkan aliran informasi di dalam *cell* dan antar *cell* maupun keluar *cell* (output h).

*Cell* LSTM mempunyai 2 hasil *output*, yang pertama adalah keluaran sebenarnya yang diteruskan lagi ke *cell* selanjutnya dan menjadi *output* dari *cell* itu dan output lainnya adalah *cell state* (C­t). Pada intinya, LSTM adalah suatu proses meng-*update* *cell state* dengan operasi-operasi *element-wise* di setiap cell LSTM.

Ke-4 *layer neuron* tersebut biasa disebut *gate*. Yang paling kiri (*layer neuron* dengan fungsi aktivasi sigmoid diperlihatkan pada gambar 8.3) adalah *forget gate*, yaitu *gate* yang menentukan apakah informasi dari *input* Xt dan *output* hi-1 diperbolehkan lewat atau tidak. Pada *output-*nya, mendekati 1 artinya “biarkan lewat” sedangkan mendekati 0 artinya “lupakan/abaikan informasi ini”. Bisa dilihat *output* dari *gate* ini akan “di adu” dengan *cell state* melalui operasi perkalian *element-wise*. Jadi,“biarkan lewat” atau “lupakan/abaikan informasi ini” mengacu pada informasi/aliran data pada *cell* *state*. Saat melewati *gate* ini maupun *gate* lainnya, input yang awalnya adalah vektor sepanjang *output* dari *cell* sebelumnya + *input* baru dari data berurut menjadi sepanjang *hidden neuron* yang dideklarasikan, sehingga operasi *element*-*wise* (bulatan warna merah muda) menjadi memungkinkan. Menurut beberapa referensi, *forget gate* adalah kunci dari kesuksesan LSTM mencegah *vanishing gradient* atau kasus sebaliknya *exploding gradient* (pembelajaran atau *update weight* menjadi terlalu cepat atau besar). [2].

## RMSprop

RMSprop adalah metode tingkat pembelajaran adaptif yang tidak dipublikasikan, metode ini diusulkan oleh Geoff Hinton.

RMSprop dan Adadelta (metode pembelajaran lainnya) telah dikembangkan secara independen pada waktu yang bersamaan yang berasal dari kebutuhan untuk menyelesaikan tingkat pembelajaran yang menurun secara radikal dari Adagrad (metode pembelajaran lainnya). Berikut fungsi derivative dari RMSprop:

E[g2]t=0.9E[g2]t−1+0.1g2t

θt+1=θt−η√E[g2]t+ϵgt

RMSprop juga membagi learning rate dengan melakukan eksponensial dari nilai rata-rata gradien. Hinton menyarankan menggunakan γ = 0.9, sementara nilai default yang bagus untuk learning rate adalah 0.001.

## *Categorical Cross-Entropy*

Cross-entropy umumnya digunakan untuk mengukur perbedaan antara dua distribusi probabilitas. Biasanya distribusi "true" (distribusi yang dicoba cocok dengan algoritma pembelajaran data) dinyatakan dalam nilai 1, dan sebaliknya “false” dinyatakan dalam nilai 0.

Sedangkan *cross entropy* dibagi menjadi dua jenis yaitu *categorical* dan *binary*. *Binary cross-entropy* digunakan hanya untuk dua *class*, sednagkan *categorical cross-entropy* digunakan untuk data yang memiliki lebih dari dua *class*.

*[Halaman ini sengaja dikosongkan]*

# BAB III ANALISIS DAN PERANCANGAN SISTEM

Bab ini akan menjelaskan tentang analisis dan perancangan sistem untuk mencapai tujuan dari tugas akhir. Perancangan ini meliputi perancangan data dan perancangan proses. Bab ini juga akan menjelaskan tentang analisis implementasi metode secara umum pada sistem.

## Analisis Metode Secara Umum

Pada tugas akhir ini akan dibangun suatu sistem untuk melakukan prediksi curah hujan di Stasiun Perak 1, Stasiun Perak 2 dan Stasiun Juanda dalam waktu perhari menggunakan pustaka R. Proses-proses yang dilakukan dalam pengimplementasian sistem ini meliputi tahap *pre-processing* atau perancangan data, tahap *processing* dimana dilakukan ekstraksi fitur, seleksi fitur dan *forecasting*, dan tahap *post-processing* atau evaluasi.

Tahap *pre-processing* atau perancangan data merupakan tahap di mana menyiapkan data yang berkaitan dengan curah hujan (berupa 8 prediktor) dari 3 stasiun, serta menyiapkan data yang berkaitan dengan el nino (berupa 4 prediktor), lalu menggabungkan pada masing-masing data stasiun yang ada.

Tahapan *processing* dilakukan dengan tiga tahap. Tahap pertama dilakukan ekstraksi fitur dengan menggunakan DPCCA. Lalu tahapan kedua yaitu melakukan seleksi prediktor yang digunakan untuk proses selanjutnya. Tahap terakhir adalah forecasting menggunakan LSTM dengan tujuan untuk mendapatkan nilai akurasi dan prediksi yang akan digunakan pada tahap evaluasi untuk dilakukan uji coba.

Tahap *post-processing* evaluasi bertujuan untuk mengetahui mana kah seleksi fitur yang paling baik untuk diimplementasikan ke dalam masing-masing skenario curah hujan. Hasil uji coba tersebut akan dibandingkan dengan metode evaluasi berupaAkurasi dan data loss *Caterogical Cross-Entropy*. Diagram alir dari keseluruhan proses ditunjukkan oleh Gambar 3.1.



Gambar 3. 1 Diagram alir implementasi metode secara umum

## Perancangan Data

Subbab ini akan membahas perancangan data yang merupakan bagian penting karena akan menjadi acuan untuk membuat skenario uji coba dari tugas akhir. Data yang digunakan adalah data curah hujan Stasiun Perak 1, Stasiun Perak 2 dan Stasiun Juanda setiap harinya dari tanggal 01/01/1987 hingga 31/12/2016.

Data historis penumpang tersebut dipisahkan sesuai dengan skenario data keberangkatan, kedatangan, transit, dan total. Masing-masing dari skenario data akan diuji coba menggunakan empat fitur, lima fitur, dan enam fitur.

Pada tahap ini, data yang didapatkan BMKG tidaklah sempurna, ada beberapa data yang tidak bernilai atau NA. Maka dari itu pada tahap ini, sebelum melakukan proses lanjut penulis akan menghilangkan prediktor yang tidak bernilai atau NA, dan menggantinya dengan beberapa cara. Jika dalam satu hari terdapat NA pada salah satu prediktor maka akan diganti dengan nilai min/maks/avg dari kurun waktu satu bulan. Misalkan, prediktor suhu minimum pada tanggal 24/07/1997 bernilai NA, maka nilai akan diganti dari nilai minimum dari prediktor suhu minimum bulan Juli tahun 1997. Setelah data sudah tidak ada yang NA, penulis menggabungkan prediktor lokal dengan prediktor global sesuai dengan kurun waktu (01/01/1987 hingga 31/12/2016). Pada tahap ini, penulis mengerjakan tahap perancangan data dengan menggunakan *tools* Ms. Excel.

### Spatio Temporal

Spatio temporal yang dimaksudkan adalah data yang digunakan berhubungan dengan tempat dan waktu. Dalam hal ini kondisi yang dibutuhkan untuk tugas akhir ini adalah jarak stasiun yang digunakan <30 km.

|  |  |  |
| --- | --- | --- |
| Stasiun 1 | Stasiun 2 | Jarak (km) |
| Perak 1 | Perak 2 | 2,00 |
| Perak 1 | Juanda | 17,79 |
| Perak 2 | Juanda | 19,49 |

Tabel 3. 1 Jarak antar stasiun

### Data *Training*

Data *training* yang digunakan adalah 80% data pertama dari data curah hujan. Berikut daftar data *training* yang digunakan:

|  |  |  |
| --- | --- | --- |
| Nama File | Jumlah Baris | Jumlah Kolom |
| opt.3perak1-linier.csv | 8.767 | 31 |
| opt.3perak2-linier.csv | 8.767 | 74 |
| opt.3juanda-linier.csv | 8.767 | 65 |
| opt.3perak1-nonlinier.csv | 8.767 | 31 |
| opt.3perak2-nonlinier.csv | 8.767 | 74 |
| opt.3juanda-nonlinier.csv | 8.767 | 65 |
| opt.3lokasi-linier.csv | 8.767 | 70 |
| opt.3lokasi-nonlinier.csv | 8.767 | 70 |

Tabel 3. 2 Detil jumlah baris dan kolom data training

### Data *Testing*

Data *testing* yang digunakan adalah 20% data terakhir dari data curah hujan. Data *testing* sekaligus digunakan sebagai validasi data. Berikut daftar data *testing* yang digunakan:

|  |  |  |
| --- | --- | --- |
| Nama File | Jumlah Baris | Jumlah Kolom |
| opt.3perak1-linier.csv | 2192 | 31 |
| opt.3perak2-linier.csv | 2192 | 74 |
| opt.3juanda-linier.csv | 2192 | 65 |
| opt.3perak1-nonlinier.csv | 2192 | 31 |
| opt.3perak2-nonlinier.csv | 2192 | 74 |
| opt.3juanda-nonlinier.csv | 2192 | 65 |
| opt.3lokasi-linier.csv | 2192 | 70 |
| opt.3lokasi-nonlinier.csv | 2192 | 70 |

Tabel 3. 3 Detil jumlah baris dan kolom data testing

## Perancangan Proses

Subbab ini membahas mengenai perancangan proses yang dilakukan untuk masing-masing metode. Hasil dari masing-masing metode ini nantinya akan dibandingkan guna mengetahuo metode terbaik pada kasus prediksi curah hujan harian.

### Metode DPCCA

Pada metode DPCCA korelasi dihitung sebagai kovarians *detrended* dari dua dataset atau lebih atas varian *detrended* dari dua seri terintegrasi, yang dalam kasus ini adalah *time series*. Dengan rumus sebagai berikut:

*xij = i*-th input variable at *j*-th location

i = 1, 2, …, 12;

j = 1, 2, 3;

*y*j  = variabel respon pada lokasi *j*-th

*s* = ukuran sel

*t* = *time lag*

T = *maximum time lag*

1. Untuk time series *yj, xij* sebagai {*xt*} digunakan untuk mengintegrasikan data:

|  |  |
| --- | --- |
|  | (2) |
| and | (3) |
|  |  |

1. Komputasi *detrended fluctuation analysis* (DFA):

|  |  |
| --- | --- |
|  | (4) |

|  |  |
| --- | --- |
|  | (5) |

1. Komputasi *detrended* kovarians of dua data:

|  |  |
| --- | --- |
|  | (6) |

Komputasi nilai rata-rata dari semua sel:

|  |  |
| --- | --- |
|  | (7) |
|  |  |

1. Komputasi nilai koefisien *detrended cross-correlation*:

|  |  |
| --- | --- |
|  | (8) |
|  |  |

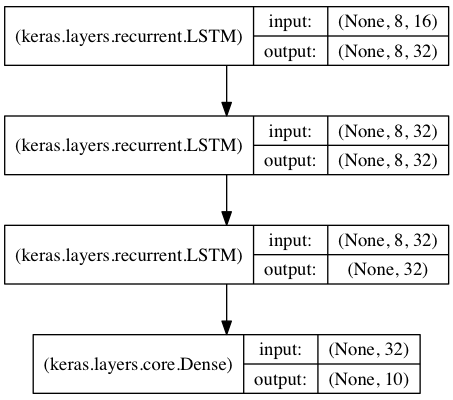
1. Komputasi koefisien DPCCA:

|  |  |
| --- | --- |
|  | (9) |

### Metode LSTM

Dalam model ini, penulis menggunakan 3 lapisan LSTM di atas satu sama lain, membuat model mampu mempelajari representasi temporal tingkat tinggi.

Dua LSTM pertama mengembalikan urutan output penuh mereka, tetapi yang terakhir hanya mengembalikan langkah terakhir dalam urutan outputnya, sehingga menjatuhkan dimensi temporal (misal mengubah urutan input menjadi vektor tunggal).



Gambar 3. 2 Stacked layer LSTM

# BAB IV IMPLEMENTASI

Bab ini membahas implementasi dari perancangan sistem sesuai dengan perancangan yang telah dibuat. Bahasa pemrograman yang digunakan untuk implementasi sistem adalah bahasa pemrograman R.

## Lingkungan Implementasi

Lingkungan implementasi sistem yang digunakan untuk mengembangkan tugas akhir ini memiliki spesifikasi perangkat keras dan perangkat lunak yang ditunjukkan oleh Tabel 4.1.

|  |  |
| --- | --- |
| **Perangkat** | **Spesifikasi** |
| Perangkat Keras | * Prosesor: Intel® Core™ i5-7500U CPU @ 2.70GHz (4 CPUs), ~2.9GHz * Memori: 8192MB |
| Perangkat Lunak | * Sistem Operasi   Microsoft Windows 10 64-bit   * Perangkat Pengembang   R Studio 64bit   * Perangkat Pembantu   Microsoft Excel 2016, Microsoft Word 2016, Microsoft Power Point 2016, Snipping Tools |

Tabel 4. 1 Tabel spesifikasi perangkat

## Implementasi Proses

Implementasi proses dilakukan berdasarkan perancangan proses yang dijelaskan pada bab analisis dan perancangan. Untuk implementasi pada semua metode, *file* yang dibaca di dalam program tergantung dari skenario data yang sedang dijalankan. Berikut adalah daftar dari nama *file* berdasarkan sebelum dilakukan seleksi fitur:

|  |  |
| --- | --- |
| **Stasiun** | **File** |
| Perak 1 | csv-harian-perak1.csv |
| Perak 2 | csv-harian-perak2.csv |
| Juanda | csv-harian-juanda.csv |

Tabel 4. 2 Tabel nama file csv sebelum dilakukan seleksi fitur

Berikut adalah daftar dari nama *file* berdasarkan hasil sesudah dilakukan seleksi fitur:

|  |  |  |
| --- | --- | --- |
| **Stasiun** | ***Linier*** | ***Non-Lininer*** |
| Perak 1 | opt.3perak1-linier.csv | opt.3perak1-nonlinier.csv |
| Perak 2 | opt.3perak2-linier.csv | opt.3perak2-nonlinier.csv |
| Juanda | opt.3juanda-linier.csv | opt.3juanda-nonlinier.csv |
| Gabungan 3 Lokasi | opt.3lokasi-linier.csv | opt.3lokasi-nonlinier.csv |

Tabel 4. 3 Tabel nama file csv sesudah dilakukan seleksi fitur

### Implementasi Metode *DPCCA*

### Implementasi Metode *Long Short Term Memories*

Dalam mengimplementasikan metode *Long Short Term Meories* (LSTM), digunakan pustaka *keras* dan *tensorflow* untuk membuat model LSTM menggunakan data *training*.Prediksi dilakukan terhadap data *testing* lalu dihitung nilai MAPE, RMSE, MAE, dan R2 dari setiap percobaan.

Fungsi yang diberi nama *SVRTEST* ini memerlukan parameter C dan ɛ sebagai masukan. Nilai ɛ yang dimasukkan sebelumnya dikalikan dengan 100.000 agar memiliki besaran yang tidak jauh berbeda dengan C saat uji coba. Pada implementasi nilai *x(1)* menunjukkan nilai C dan nilai *x(2)* menunjukkan nilai ɛ. Jumlah dan nama fitur yang ada di dalam fungsi bergantung dari jumlah fitur yang sedang diuji coba. Nama *kernel* diubah-ubah sesuai dengan *kernel* yang sedang digunakan dalam uji coba. Kode Sumber 4.3, Kode Sumber 4.4, dan Kode Sumber 4.5 menunjukkan pengimplementasian dari metode SVR.

|  |  |
| --- | --- |
| 01  02  03  04  05  06  07  08  09  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95  96  97  98  99  100  101  102  103  104  105  106  107  108  109  110  111  112  113  114  115  116 | library**(**tibble**)**  library**(**keras**)**  library**(**tensorflow**)**  library**(**ggplot2**)**  # prepare data  setwd**(**"D:/Kuliah/TA/workspace"**)**  load**(**"D:/Kuliah/TA/workspace/a\_lstm\_barubaru.RData"**)**  data **<-** read.csv**(**"opt.3perak1-linier-2.csv", header **=** **TRUE**, sep **=** ","**)**  data **<-** read.csv**(**"opt.3perak2-linier-2.csv", header **=** **TRUE**, sep **=** ","**)**  data **<-** read.csv**(**"opt.3juanda-linier-2.csv", header **=** **TRUE**, sep **=** ","**)**  data**[**is.na**(**data**)]** **<-** 0  data **<-** data.matrix**(**data**[**,**-**1**])**  target **<-** read.csv**(**"csv-harian-perak1.csv", header **=** **TRUE**, sep **=** ","**)**  target **<-** read.csv**(**"csv-harian-perak2.csv", header **=** **TRUE**, sep **=** ","**)**  target **<-** read.csv**(**"csv-harian-juanda.csv", header **=** **TRUE**, sep **=** ","**)**  target **<-** data.matrix**(**target**[**,8**])**  target**[**is.na**(**target**)]** **<-** 0  data **<-** cbind**(**data,target**)**  head**(**data**)**  N **=** nrow**(**data**)**  p **=** ncol**(**data**)**  # bagi banyak training testing  n **=** round**(**N **\***0.8, digits **=** 0**)**  m **=** N **-** n  # normalisasi  data **<-** **(**data **-** min**(**data**))** **/** **(**max**(**data**)** **-** min**(**data**))**  x\_train **=** data.matrix**(**data**[**1**:**n, 1**:**p**])**  y\_train **<-** array\_reshape**(**x **=** x\_train, dim **=** list**(**n, 1, p**))**  x\_test **=** data.matrix**(**data**[(**n**+**1**):**N, **])**  y\_test **<-** array\_reshape**(**x **=** x\_test, dim **=** list**(**m, 1, p**))**  # constants  units **=** 32  timesteps **=** 1  # define and compile model  # expected input data shape: (batch\_size, timesteps, feature)  # stacked lstm sequence  model **<-** keras\_model\_sequential**()**  model %>%  layer\_lstm**(**units, return\_sequences **=** **TRUE**, input\_shape **=** c**(**timesteps, p**))** %>%  layer\_lstm**(**units, return\_sequences **=** **TRUE)** %>%  layer\_lstm**(**units**)** %>%  layer\_dense**(**units, activation **=** 'softmax'**)**  # compile  rm**(**model**)**  # 1. rmsprop - categorical  model %>% compile**(**  loss **=** 'categorical\_crossentropy',  optimizer **=** 'rmsprop',  metrics **=** c**(**'accuracy'**)**  **)**  # 2. adam - mse  model %>% compile**(**  loss **=** 'mean\_squared\_error',  optimizer **=** optimizer\_adam**(** lr**=** 0.02, decay **=** 1e**-**6 **)**,  metrics **=** c**(**'accuracy'**)**  **)**  # 3. rmsprop - mae  model %>% compile**(**  loss **=** "mae",  optimizer **=** optimizer\_rmsprop**()**,  metrics **=** c**(**'accuracy'**)**  **)**  # summary  summary**(**model**)**  # train  history **<-** model %>% fit**(**  y\_train,  x\_train,  batch\_size **=** units,  epochs **=** 1000,  validation\_data **=** list**(**y\_test, x\_test**)**  #validation\_split = 0.2  **)**  # evaluate  model %>% evaluate**(**x\_train, y\_train**)**  model %>% evaluate**(**x\_test, y\_test**)**  # predict  model %>% predict\_classes**(**x\_test**)**  # evaluate the model  evals **<-** model %>% evaluate**(**x\_test, y\_test, batch\_size **=** 10**)**  accuracy **=** evals**[**2**][[**1**]]\*** 100  # check accuracy  accuracy |

# BAB V PENGUJIAN DAN EVALUASI

Bab ini membahas uji coba dan evaluasi terhadap sistem yang telah dikembangkan untuk memprediksi jumlah penumpang Bandar Udara Juanda perbulan dari data historis penumpang bulan Januari 2010 hingga Desember 2015.

## Lingkungan Pengujian

Lingkungan penguijan sistem pada pengerjaan tugas ini dilakukan pada lingkungan dan alat kakas sebagai berikut:

Prosesor : Prosesor: Intel® Core™ i5-7500U CPU @ 2.70GHz (4 CPUs) ~2.9GHz

RAM : 8192 MB

Jenis *Device* : Laptop

Sistem Operasi : Microsoft Windows 10 64-bit

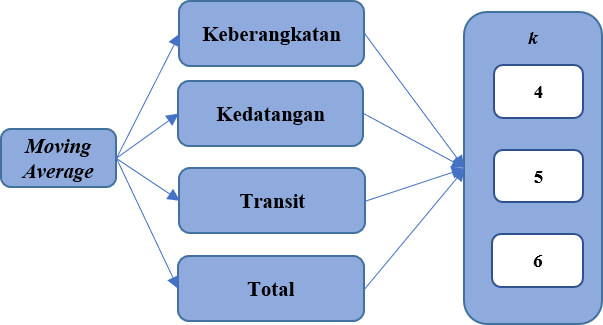
## Data Uji Coba

Data yang digunakan untuk uji coba prediksi jumlah penumpang Bandar Udara Juanda adalah data historis penumpang perbulan yang merupakan 10% dari keseluruhan data yang ada untuk masing-masing skenario. Pembagian data *training* dan data *testing* sebelumnya telah dijelaskan pada bagian Perancangan Data. Jumlah data *testing* yang digunakan berjumlah 18 atau 19 baris.

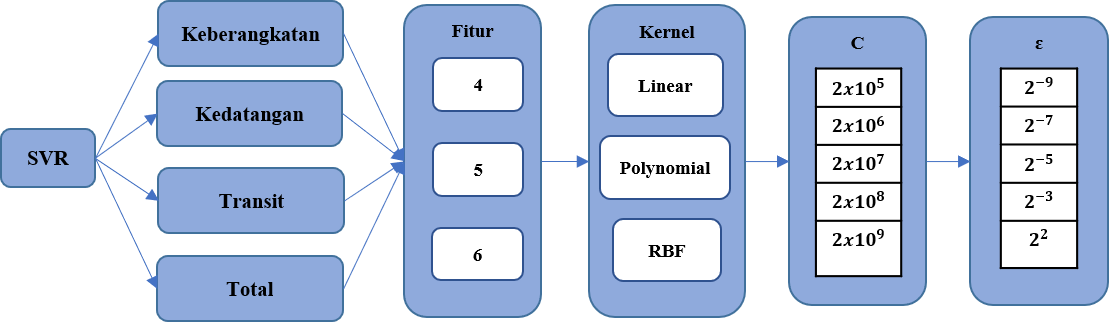
## Skenario Uji Coba

Subbab ini akan menjelaskan skenario uji coba yang telah dilakukan. Terdapat beberapa skenario uji coba yang telah dilakukan. Setiap skenario uji coba yang berdasarkan metode ini memiliki beberapa sub skenario yang dapat dilihat di masing-masing skenario uji coba.

1. Skenario Pengujian 1: perhitungan nilai *Mean Absolute Percentage Error* (MAPE), *Root Mean-Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi (R2)pada pengujian dengan metode *Moving Average*. Terdapat beberapa skenario percobaan dengan metode ini yang dapat dilihat pada Gambar 5.1.



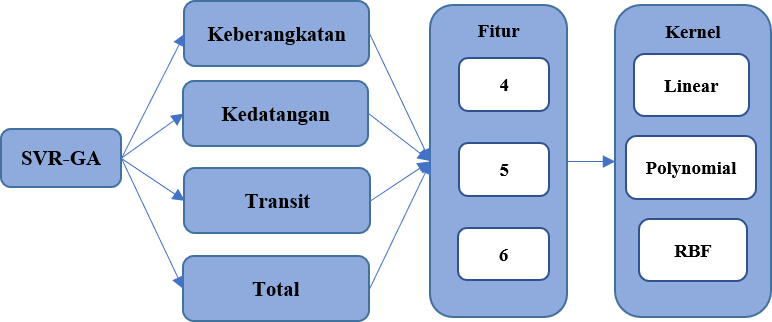
Gambar 5. 1 Skenario Uji Coba Moving Average

1. Skenario Pengujian 2: perhitungan nilai *Mean Absolute Percentage Error* (MAPE), *Root Mean-Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi (R2) pada pengujian dengan metode *Support Vector Regression* (SVR). Terdapat beberapa skenario percobaan dengan metode ini yang dapat dilihat pada Gambar 5.2.

Gambar 5. 2 Skenario Uji Coba SVR

Pada skenario ini, masing-masing data akan diuji coba menggunakan tiga macam *kernel*. Setiap *kernel* akan diuji coba menggunakan empat, lima, dan enam fitur. Setiap percobaan tersebut mengkombinasikan C dan ɛ untuk menemukan model terbaik.

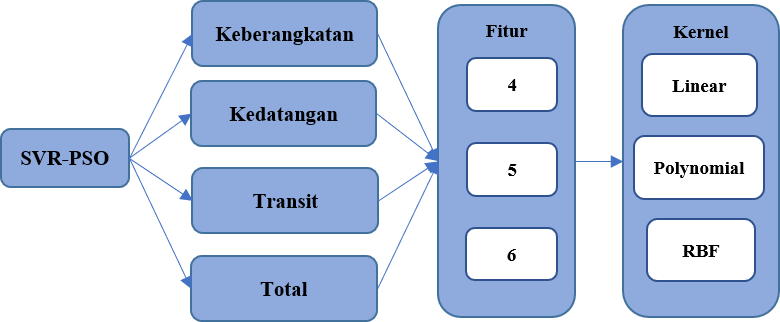
1. Skenario Pengujian 3: perhitungan nilai *Mean Absolute Percentage Error* (MAPE), *Root Mean-Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi (R2) pada pengujian dengan metode *Support Vector Regression – Genetic Algorithm* (SVR-GA). Terdapat beberapa skenario percobaan dengan metode ini yang dapat dilihat pada Gambar 5.3.



Gambar 5. 3 Skenario Uji Coba SVR-GA

Pada skenario ini, setiap data diuji coba menggunakan tiga macam *kernel* ke dalam empat, lima, dan enam fitur. Jumlah generasi maksimal yang digunakan di dalam algoritma *Genetic Algorithm* adalah sebanyak 60 generasi dengan jumlah populasi sebesar 50. *Selection function* yang digunakan adalah *the Tournament*. Rentang nilai C yang digunakan adalah 200000 – 20000000 dan nilai ɛ adalah 0,002 – 4.

1. Skenario Pengujian 4: perhitungan nilai *Mean Absolute Percentage Error* (MAPE), *Root Mean-Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan koefisien determinasi (R2) pada pengujian dengan metode *Support Vector Regression – Particle Swarm Optimization* (SVR-PSO). Terdapat beberapa skenario percobaan dengan metode ini yang dapat dilihat pada Gambar 5.4.



Gambar 5. 4 Skenario Uji Coba SVR-PSO

Setiap data di dalam skenario ini diuji coba menggunakan tiga macam *kernel* ke dalam empat, liat, dan enam fitur. Jumlah iterasi maksimal yang digunakan di dalam algoritma *Particle Swarm Optimization* adalah sebanyak 400 iterasi dengan jumlah partikel sebanyak 20. Rentang nilai C yang digunakan adalah 200000 – 20000000 dan nilai ɛ adalah 0,002 – 4.

## Skenario Pengujian 1

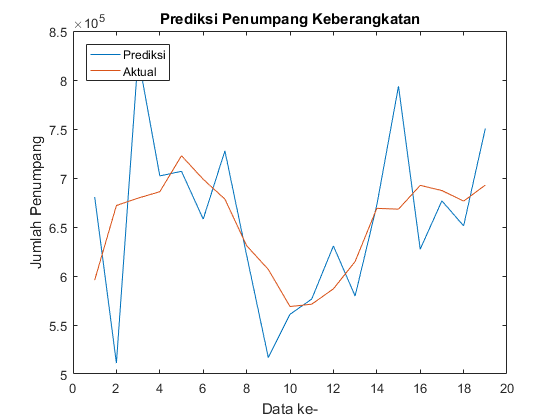
Pada skenario ini dilakukan uji coba prediksi data penumpang Bandar Udara Juanda perbulan menggunakan algoritma *Moving Average* (MA) di setiap masing-masing data menggunakan *k* = {4, 5, 6}. Selanjutnya dilakukan proses perhitungan MAPE, RMSE, MAE, dan R2.

### Skenario Pengujian 1.1

Bagian ini akan menjelaskan hasil dari uji coba menggunakan data Keberangkatan menggunakan algoritma *Moving Average*. Berikut adalah hasil uji coba yang telah dilakukan menggunakan data *training*.

Tabel 5. 1 Hasil Uji Coba Data Training Keberangkatan (MA)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***k*** | **MAPE (%)** | **RMSE** | **MAE** | **R2** |
| 4 | 6,426 | 30922,281 | 21493,989 | 0,968 |
| 5 | 7,062 | 32271,522 | 23278,465 | 0,965 |
| 6 | 7,442 | 33385,001 | 24722,294 | 0,963 |

Hasil uji coba terbaik ditunjukkan saat menggunakan nilai *k =* 4 pada data *training*. Selanjutnya nilai *k* = 4 digunakan ke dalam data *testing* sehingga menghasilkan nilai MAPE sebesar 8,052%; RMSE sebesar 70328,36; MAE sebesar 52148,127; dan R2 sebesar 0,285. Berikut adalah *plotting* dari uji coba menggunakan data *testing*.

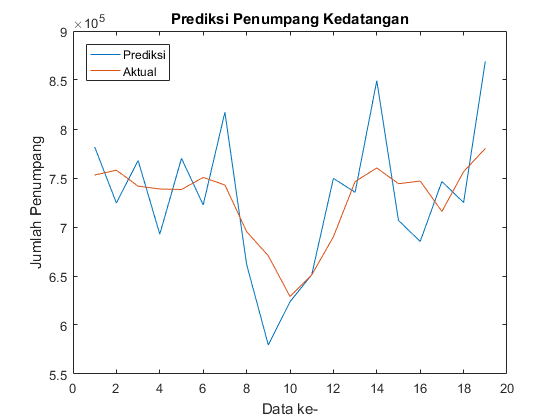
Gambar 5. 5 Plot Data Testing Keberangkatan (MA)

### Skenario Pengujian 1.2

Bagian ini akan menjelaskan hasil dari uji coba menggunakan data Kedatangan menggunakan algoritma *Moving Average*. Berikut adalah hasil uji coba yang telah dilakukan menggunakan data *training*.

Tabel 5. 2 Hasil Uji Coba Data Training Kedatangan (MA)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***k*** | **MAPE (%)** | **RMSE** | **MAE** | **R2** |
| 4 | 5,523 | 25058,159 | 18795,079 | 0,982 |
| 5 | 5,830 | 25227,646 | 19676,148 | 0,982 |
| 6 | 6,009 | 27569,701 | 21145,317 | 0,979 |

Hasil uji coba terbaik ditunjukkan saat menggunakan nilai *k =* 4 pada data *training*. Selanjutnya nilai *k* = 4 digunakan ke dalam data *testing* sehingga menghasilkan nilai MAPE sebesar 5,787%; RMSE sebesar 50316,800; MAE sebesar 42526,202; dan R2 sebesar 0,501. Berikut adalah *plotting* dari uji coba menggunakan data *testing*.

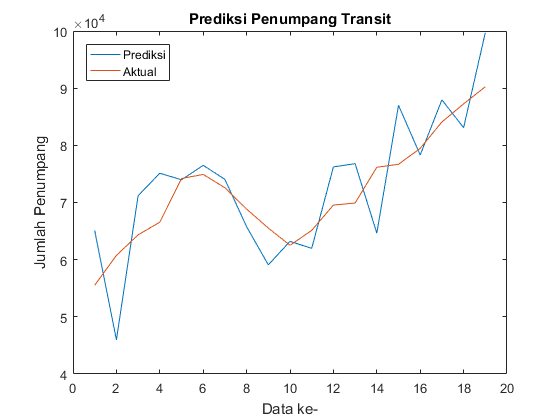
Gambar 5. 6 Plot Data Testing Kedatangan (MA)

### Skenario Pengujian 1.3

Bagian ini akan menjelaskan hasil dari uji coba menggunakan data Transit menggunakan algoritma *Moving Average*. Berikut adalah hasil uji coba yang telah dilakukan menggunakan data *training*.

Tabel 5. 3 Hasil Uji Coba Data Training Transit (MA)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***k*** | **MAPE (%)** | **RMSE** | **MAE** | **R2** |
| 4 | 10,871 | 9713,450 | 5365,037 | 0,779 |
| 5 | 11,182 | 9856,875 | 5462,325 | 0,772 |
| 6 | 11,833 | 10193,265 | 5829,084 | 0,757 |

Hasil uji coba terbaik ditunjukkan saat menggunakan nilai *k =* 4 pada data *training*. Selanjutnya nilai *k* = 4 digunakan ke dalam data *testing* sehingga menghasilkan nilai MAPE sebesar 8,518 %; RMSE sebesar 7074,759; MAE sebesar 5815,206; dan R2 sebesar 0,639; Berikut adalah *plotting* dari uji coba menggunakan data *testing*.

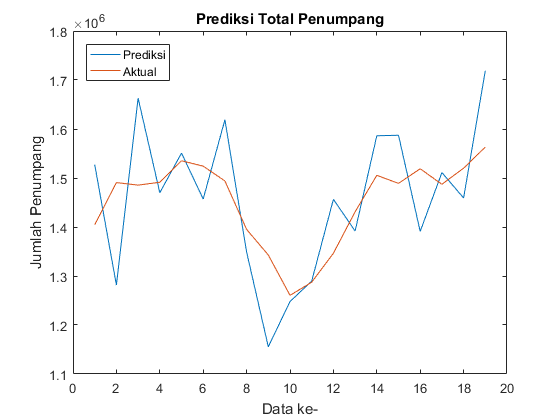
Gambar 5. 7 Plot Data Testing Transit (MA)

### Skenario Pengujian 1.4

Bagian ini akan menjelaskan hasil dari uji coba menggunakan data Total menggunakan algoritma *Moving Average*. Berikut adalah hasil uji coba yang telah dilakukan menggunakan data *training*.

Tabel 5. 4 Hasil Uji Coba Data Training Total (MA)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***k*** | **MAPE (%)** | **RMSE** | **MAE** | **R2** |
| 4 | 5,338 | 53302,323 | 38278,472 | 0,980 |
| 5 | 5,559 | 53846,714 | 39935,807 | 0,979 |
| 6 | 5,835 | 57488,920 | 43187,741 | 0,976 |

Hasil uji coba terbaik ditunjukkan saat menggunakan nilai *k =* 4 pada data *training*. Selanjutnya nilai *k* = 4 digunakan ke dalam data *testing* sehingga menghasilkan nilai MAPE sebesar 6,121%; RMSE sebesar 108401,788; MAE sebesar 88614,509; dan R2 sebesar 0,444. Berikut adalah *plotting* dari uji coba menggunakan data *testing*.

Gambar 5. 8 Plot Data Testing Total (MA)

## Skenario Pengujian 2

Pada skenario ini dilakukan uji coba prediksi data penumpang Bandar Udara Juanda perbulan menggunakan algoritma *Support Vector Regression* (SVR) di setiap masing-masing data yaitu data keberangkatan, kedatangan, transit, dan total. Masing-masing skenario data akan diuji coba menggunakan tiga jenis *kernel* yaitu linear, polynomial, dan *Radial Basis Function* (RBF) ke dalam tiga fitur dengan mengkombinasikan berbagai C, dan ɛ yang berbeda-beda. Selanjutnya dilakukan proses perhitungan MAPE, RMSE, MAE, dan R2 sebagai evaluasi dan pembanding untuk mengetahui hasil uji coba terbaik.

### Skenario Pengujian 2.1

Bagian ini akan menjelaskan hasil dari uji coba menggunakan data Keberangkatan menggunakan algoritma *Support Vector Regression* (SVR) dengan menerapkan berbagai fungsi *kernel*, fitur, dan berbagai kombinasi nilai C dan ɛ

.

* 1. **Skenario Pengujian 2.1.1**

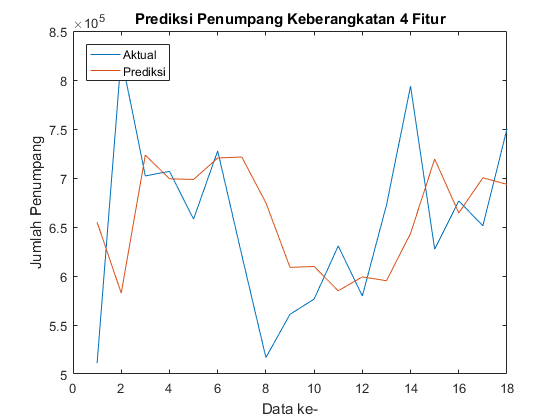
Bagian ini menjelaskan hasil uji coba pada data Keberangkatan dengan algoritma SVR menggunakan empat fitur.

1. *Kernel* Linear

Tabel 5.5 menunjukkan hasil uji coba menggunakan *kernel* linear dengan berbagai kombinasi nilai C dan ɛ. Berdasarkan tabel tersebut, hasil terbaik ditunjukkan saat nilai C sama dengan 2x105 dan nilai ɛ sama dengan 2-7 dengan MAPE sebesar 11,23713%; RMSE sebesar 95670,83; MAE sebesar 72431,28; dan R2 sebesar -0,2591. *Plotting* dari hasil uji coba dengan nilai MAPE terbaik pada skenario ini ditunjukkan oleh Gambar 5.9.

Tabel 5. 5 Hasil Uji Coba Data Keberangkatan 4 Fitur Kernel Linear (SVR)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **C** | ɛ | **MAPE (%)** | **RMSE** | **MAE** | **R2** |
| 2x105 | 2-9 | 11,24508 | 95687,00 | 72490,28 | -0,2595 |
| 2-7 | 11,23713 | 95670,83 | 72431,28 | -0,2591 |
| 2-5 | 11,23752 | 95670,99 | 72432,45 | -0,2591 |
| 2-3 | 11,24458 | 95684,53 | 72487,31 | -0,2595 |
| 22 | 11,23936 | 95683,17 | 72451,98 | -0,2594 |
| 2x106 | 2-9 | 11,23921 | 95728,21 | 72511,85 | -0,2606 |
| 2-7 | 11,24064 | 95732,55 | 72520,50 | -0,2607 |
| 2-5 | 11,24135 | 95733,09 | 72525,67 | -0,2607 |
| 2-3 | 11,24041 | 95730,70 | 72518,00 | -0,2607 |
| 22 | 11,24018 | 95730,94 | 72516,99 | -0,2607 |
| 2x107 | 2-9 | 11,23912 | 95727,65 | 72510,33 | -0,2606 |
| 2-7 | 11,24107 | 95733,68 | 72524,38 | -0,2608 |
| 2-5 | 11,24148 | 95734,25 | 72527,49 | -0,2608 |
| 2-3 | 11,24033 | 95730,70 | 72517,68 | -0,2607 |
| 22 | 11,23925 | 95728,42 | 72512,08 | -0,2606 |
| 2x108 | 2-9 | 83,18935 | 561332,60 | 528107,20 | -42,3453 |
| 2-7 | 83,18923 | 561331,90 | 528106,40 | -42,3452 |
| 2-5 | 83,18874 | 561328,90 | 528103,30 | -42,3447 |
| 2-3 | 83,18678 | 561316,90 | 528090,70 | -42,3429 |
| 22 | 83,10579 | 560820,40 | 527569,20 | -42,2662 |
| 2x109 | 2-9 | 37,22449 | 273674,60 | 233138,20 | -9,3032 |
| 2-7 | 37,22170 | 273655,70 | 233120,50 | -9,3017 |
| 2-5 | 49,10058 | 341960,10 | 307205,70 | -15,0862 |
| 2-3 | 48,99686 | 341271,80 | 306533,10 | -15,0215 |
| 22 | 123,24120 | 813612,90 | 784633,00 | -90,0620 |



Gambar 5. 9 Plot Terbaik Data Keberangkatan 4 Fitur Kernel Linear (SVR)

# BAB VI KESIMPULAN DAN SARAN

Pada bab ini akan diberikan kesimpulan yang diperoleh selama pengerjaan tugas akhir dan saran mengenai pengembangan yang dapat dilakukan terhadap tugas akhir ini di masa yang akan datang

## Kesimpulan

Dari hasil pengamatan selama proses perancangan, implementasi, dan pengujian yang dilakukan, dapat diambil kesimpulan sebagai berikut.

* 1. Pada data keberangkatan, MAPE terbaik didapatkan sebesar 6,6696% menggunakan metode SVR-PSO dengan 6 fitur, *kernel* RBF, nilai C sama dengan 1035237,51443656; dan nilai ɛ sama dengan 3,82943088639146.
  2. Pada data kedatangan, MAPE terbaik didapatkan sebesar 5,7870% menggunakan metode *Moving Average* dengan nilai *k* sama dengan 4. Disusul oleh MAPE sebesar 7,3784% menggunakan metode SVR-PSO dengan 6 fitur, *kernel* linear, nilai C sama dengan 1379172,86163654; dan nilai ɛ sama dengan 3,36573392951344.
  3. Pada data transit, MAPE terbaik didapatkan sebesar 8,5180% menggunakan metode *Moving Average* dengan nilai *k* sama denagn 4. Disusul oleh MAPE sebesar 11,6187% menggunakan metode SVR-PSO dengan 6 fitur, *kernel* linear, nilai C sama dengan 16184927,0061892; dan nilai ɛ sama dengan 2,66865236867938.
  4. Pada data total, MAPE terbaik didapatkan sebesar 6,1210% menggunakan metode *Moving Average* dengan nilai *k* sama dengan 4. Disusul oeh MAPE sebesar 6,2559% menggunakan metode SVR-PSO dengan 6 fitur, *kernel* RBF, nilai C sama dengan 1386963,2508749, dan nilai ɛ sama dengan 3,1126895667879.
  5. Metode SVR-PSO terbukti lebih baik dibandingkan dengan metode SVR atau SVR-GA pada kasus prediksi penumpang Bandar Udara Juanda perbulan.
  6. Pada data kedatangan, transit, dan total, penggunaan metode sederhana yaitu *Moving Average* lebih baik dari metode SVR, SVR-GA, dan SVR-PSO.

## Saran

Berikut merupakan beberapa saran untuk pengembangan sistem di masa yang akan dating. Saran-saran ini didasarkan pada hasil perancangan, implementasi, dan pengujian yang telah dilakukan.

1. Selain menggunakan empat, lima, dan enam fitur, lakukan percobaan menggunakan jumlah fitur yang berbeda dengan memperhitungkan korelasi dari bulan-bulan lainnya untuk dijadikan fitur.
2. Tambahkan fitur yang berupa faktor eksternal jika memungkinkan seperti hari libur yang ada di setiap bulan serta kebijakan pemerintah yang ada.
3. Menerapkan algoritma regresi selain SVR atau *Moving Average* untuk meningkatkan akurasi
4. Melakukan *pre-processing* untuk mengolah data supaya dapat menghasilkan prediksi yang lebih akurat.

# DAFTAR PUSTAKA

|  |  |
| --- | --- |
| [1] | X. Zhang, P. Wang, D. Liang, C. Fan and C. Li, "A soft self-repairing for FBG sensor network in SHM system based on PSO-SBR model reconstruction," *Elsevier,* pp. 38-46, 2014. |
| [2] | K. Cheng, Z. Lu, Y. Wei, Y. Shi and Y. Zhou, "Mixed Kernel Function Support Vector Regression for Global Sensitivity Analysis," *Elsevier,* vol. 96, pp. 201-214, 2017. |
| [3] | "Support Vector Machine Regression," [Online]. Available: http://kernelsvm.tripod.com/. [Accessed 18 Mei 2017]. |
| [4] | T. Qin, S. Zeng and J. Guo, "Robust prognostics for state of health estimation of lithium-ion batteries based on an improved PSO-SVR model," *Microelectronics Reliability,* vol. 55, pp. 1280-1284, 2015. |
| [5] | B. Santosa and P. Willy, Metoda Metaheuristik, Konsep dan Implementasi, Surabaya: Guna Widya, 2011. |
| [6] | X. Wang, J. Wen, Y. Zhang and Y. Wang, "Real estate price forecasting based on SVM optimized by PSO," *Optik,* vol. 125, pp. 1439-1443, 2014. |
| [7] | T. Afonja, "Medium," 1 Januari 2017. [Online]. Available: https://towardsdatascience.com/kernel-function-6f1d2be6091. [Accessed 3 Desember 2017]. |

*[Halaman ini sengaja dikosongkan]*

# LAMPIRAN

Data Kedatangan, Keberangkatan, Transit, dan Total.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tahun** | **Bulan** | **Kedatangan** | **Keberangkatan** | **Transit** | **Total** |
| 2000 | 1 | 108704 | 106832 | 11846 | 227382 |
| 2000 | 2 | 87538 | 107287 | 10632 | 205457 |
| 2000 | 3 | 108387 | 106980 | 12276 | 227643 |
| 2000 | 4 | 111586 | 91193 | 11641 | 214420 |
| 2000 | 5 | 87289 | 83592 | 12138 | 183019 |
| 2000 | 6 | 104359 | 95569 | 12027 | 211955 |
| 2000 | 7 | 128716 | 123838 | 13902 | 266456 |
| 2000 | 8 | 99144 | 94349 | 12412 | 205905 |
| 2000 | 9 | 102667 | 99027 | 15137 | 216831 |
| 2000 | 10 | 113148 | 104864 | 15042 | 233054 |
| 2000 | 11 | 118250 | 106238 | 15225 | 239713 |
| 2000 | 12 | 127413 | 105035 | 15557 | 248005 |
| 2001 | 1 | 112797 | 135546 | 15974 | 264317 |
| 2001 | 2 | 99857 | 141977 | 14265 | 256099 |
| 2001 | 3 | 149872 | 124512 | 17830 | 292214 |
| 2001 | 4 | 128506 | 107016 | 17231 | 252753 |
| 2001 | 5 | 109330 | 106403 | 15859 | 231592 |
| 2001 | 6 | 124547 | 118026 | 16532 | 259105 |
| 2001 | 7 | 143536 | 140468 | 17778 | 301782 |
| 2001 | 8 | 127933 | 124721 | 16894 | 269548 |
| 2001 | 9 | 122413 | 122313 | 15744 | 260470 |
| 2001 | 10 | 146747 | 138773 | 18309 | 303829 |
| 2001 | 11 | 135075 | 121925 | 16157 | 273157 |
| 2001 | 12 | 165399 | 153756 | 17330 | 336485 |
| 2002 | 1 | 139664 | 170751 | 19720 | 330135 |
| **Tahun** | **Bulan** | **Kedatangan** | **Keberangkatan** | **Transit** | **Total** |
| 2002 | 2 | 121042 | 135043 | 17056 | 273141 |
| 2002 | 3 | 195232 | 158479 | 27506 | 381217 |
| 2002 | 4 | 150056 | 144243 | 35261 | 329560 |
| 2002 | 5 | 184724 | 159177 | 39023 | 382924 |
| 2002 | 6 | 191190 | 172882 | 48041 | 412113 |
| 2002 | 7 | 224983 | 204963 | 40238 | 470184 |
| 2002 | 8 | 200939 | 167664 | 47992 | 416595 |
| 2002 | 9 | 180429 | 165986 | 54303 | 400718 |
| 2002 | 10 | 211709 | 190662 | 57492 | 459863 |
| 2002 | 11 | 187634 | 157830 | 42021 | 387485 |
| 2002 | 12 | 229573 | 221479 | 50782 | 501834 |
| 2003 | 1 | 216847 | 250843 | 49869 | 517559 |
| 2003 | 2 | 213253 | 179207 | 43074 | 435534 |
| 2003 | 3 | 227403 | 198954 | 55251 | 481608 |
| 2003 | 4 | 197827 | 183897 | 50641 | 432365 |
| 2003 | 5 | 225393 | 205270 | 57315 | 487978 |
| 2003 | 6 | 248510 | 237086 | 57286 | 542882 |
| 2003 | 7 | 273648 | 261190 | 70759 | 605597 |
| 2003 | 8 | 250874 | 243215 | 66444 | 560533 |
| 2003 | 9 | 260741 | 244258 | 76888 | 581887 |
| 2003 | 10 | 284741 | 251230 | 83515 | 619486 |
| 2003 | 11 | 273297 | 207469 | 54120 | 534886 |
| 2003 | 12 | 304111 | 321627 | 155687 | 781425 |
| 2004 | 1 | 307528 | 315039 | 56870 | 679437 |
| 2004 | 2 | 326790 | 284286 | 50928 | 662004 |
| 2004 | 3 | 302925 | 274042 | 47258 | 624225 |
| 2004 | 4 | 306566 | 273888 | 50818 | 631272 |
| 2004 | 5 | 349080 | 317403 | 43058 | 709541 |
| 2004 | 6 | 361745 | 331999 | 41974 | 735718 |
| **Tahun** | **Bulan** | **Kedatangan** | **Keberangkatan** | **Transit** | **Total** |
| 2004 | 7 | 386717 | 360371 | 38978 | 786066 |
| 2004 | 8 | 384067 | 348439 | 45499 | 778005 |
| 2004 | 9 | 370927 | 334669 | 40354 | 745950 |
| 2004 | 10 | 357183 | 301397 | 37596 | 696176 |
| 2004 | 11 | 398528 | 340415 | 38600 | 777543 |
| 2004 | 12 | 350904 | 343562 | 42344 | 736810 |
| 2005 | 1 | 355844 | 350210 | 46095 | 752149 |
| 2005 | 2 | 335697 | 293192 | 38972 | 667861 |
| 2005 | 3 | 321859 | 292316 | 41002 | 655177 |
| 2005 | 4 | 306123 | 289268 | 39604 | 634995 |
| 2005 | 5 | 321873 | 308686 | 33691 | 664250 |
| 2005 | 6 | 314738 | 292644 | 23751 | 631133 |
| 2005 | 7 | 375291 | 351494 | 29856 | 756641 |
| 2005 | 8 | 353768 | 324413 | 46253 | 724434 |
| 2005 | 9 | 356187 | 328830 | 40615 | 725632 |
| 2005 | 10 | 314059 | 244057 | 29118 | 587234 |
| 2005 | 11 | 343579 | 340895 | 31740 | 716214 |
| 2005 | 12 | 350018 | 365187 | 35445 | 750650 |
| 2006 | 1 | 384260 | 347192 | 30112 | 761564 |
| 2006 | 2 | 319676 | 293704 | 25792 | 639172 |
| 2006 | 3 | 345565 | 317779 | 32185 | 695529 |
| 2006 | 4 | 350576 | 319511 | 31449 | 701536 |
| 2006 | 5 | 360194 | 329829 | 29300 | 719323 |
| 2006 | 6 | 374486 | 346280 | 35098 | 755864 |
| 2006 | 7 | 411197 | 390893 | 48556 | 850646 |
| 2006 | 8 | 393406 | 358129 | 50320 | 801855 |
| 2006 | 9 | 371549 | 335687 | 50144 | 757380 |
| 2006 | 10 | 374503 | 299548 | 35883 | 709934 |
| 2006 | 11 | 359556 | 379872 | 39462 | 778890 |
| **Tahun** | **Bulan** | **Kedatangan** | **Keberangkatan** | **Transit** | **Total** |
| 2006 | 12 | 377166 | 387542 | 50249 | 814957 |
| 2007 | 1 | 370928 | 336048 | 39490 | 746466 |
| 2007 | 2 | 278325 | 263700 | 34833 | 576858 |
| 2007 | 3 | 326696 | 302327 | 40263 | 669286 |
| 2007 | 4 | 327668 | 299900 | 40559 | 668127 |
| 2007 | 5 | 349438 | 318738 | 41630 | 709806 |
| 2007 | 6 | 351666 | 325213 | 41887 | 718766 |
| 2007 | 7 | 408235 | 375627 | 49328 | 833190 |
| 2007 | 8 | 401923 | 355864 | 44321 | 802108 |
| 2007 | 9 | 353847 | 298766 | 44162 | 696775 |
| 2007 | 10 | 402088 | 349227 | 37074 | 788389 |
| 2007 | 11 | 345347 | 361920 | 46574 | 753841 |
| 2007 | 12 | 417886 | 390879 | 50851 | 859616 |
| 2008 | 1 | 404325 | 364415 | 50494 | 819234 |
| 2008 | 2 | 345039 | 312028 | 32798 | 689865 |
| 2008 | 3 | 384474 | 354915 | 44217 | 783606 |
| 2008 | 4 | 349934 | 317167 | 45403 | 712504 |
| 2008 | 5 | 344463 | 319149 | 53564 | 717176 |
| 2008 | 6 | 390253 | 354443 | 61251 | 805947 |
| 2008 | 7 | 389614 | 351374 | 57178 | 798166 |
| 2008 | 8 | 363586 | 332256 | 55524 | 751366 |
| 2008 | 9 | 338751 | 258542 | 34429 | 631722 |
| 2008 | 10 | 374046 | 389660 | 41984 | 805690 |
| 2008 | 11 | 350550 | 373347 | 48637 | 772534 |
| 2008 | 12 | 419077 | 371739 | 43570 | 834386 |
| 2009 | 1 | 408167 | 368817 | 46301 | 823285 |
| 2009 | 2 | 353913 | 324032 | 40417 | 718362 |
| 2009 | 3 | 398550 | 358372 | 55030 | 811952 |
| 2009 | 4 | 372638 | 346921 | 56254 | 775813 |
| **Tahun** | **Bulan** | **Kedatangan** | **Keberangkatan** | **Transit** | **Total** |
| 2009 | 5 | 405933 | 379708 | 55653 | 841294 |
| 2009 | 6 | 443398 | 400027 | 66952 | 910377 |
| 2009 | 7 | 469496 | 431139 | 64862 | 965497 |
| 2009 | 8 | 445015 | 381940 | 61222 | 888177 |
| 2009 | 9 | 462012 | 360146 | 40381 | 862539 |
| 2009 | 10 | 435642 | 504834 | 50909 | 991385 |
| 2009 | 11 | 452190 | 455120 | 51486 | 958796 |
| 2009 | 12 | 501183 | 472578 | 41668 | 1015429 |
| 2010 | 1 | 434273 | 434799 | 44638 | 913710 |
| 2010 | 2 | 416607 | 385242 | 47948 | 849797 |
| 2010 | 3 | 416673 | 382365 | 57711 | 856749 |
| 2010 | 4 | 441949 | 445062 | 35877 | 922888 |
| 2010 | 5 | 487550 | 462173 | 42242 | 991965 |
| 2010 | 6 | 503788 | 489687 | 39253 | 1032728 |
| 2010 | 7 | 549155 | 521390 | 52067 | 1122612 |
| 2010 | 8 | 483088 | 405718 | 37498 | 926304 |
| 2010 | 9 | 533657 | 510548 | 25479 | 1069684 |
| 2010 | 10 | 498405 | 552465 | 60712 | 1111582 |
| 2010 | 11 | 536705 | 519832 | 62291 | 1118828 |
| 2010 | 12 | 548052 | 535999 | 71161 | 1155212 |
| 2011 | 1 | 522281 | 515682 | 54182 | 1092145 |
| 2011 | 2 | 476818 | 457739 | 55118 | 989675 |
| 2011 | 3 | 537040 | 500819 | 55560 | 1093419 |
| 2011 | 4 | 498959 | 474766 | 58543 | 1032268 |
| 2011 | 5 | 529245 | 498259 | 59996 | 1087500 |
| 2011 | 6 | 557883 | 523916 | 60481 | 1142280 |
| 2011 | 7 | 619268 | 561883 | 71050 | 1252201 |
| 2011 | 8 | 543295 | 397653 | 38760 | 979708 |
| 2011 | 9 | 578084 | 610246 | 124642 | 1312972 |
| **Tahun** | **Bulan** | **Kedatangan** | **Keberangkatan** | **Transit** | **Total** |
| 2011 | 10 | 560334 | 603303 | 67051 | 1230688 |
| 2011 | 11 | 627030 | 569815 | 68532 | 1265377 |
| 2011 | 12 | 636694 | 584780 | 72134 | 1293608 |
| 2012 | 1 | 656251 | 613125 | 59577 | 1328953 |
| 2012 | 2 | 573264 | 546230 | 54839 | 1174333 |
| 2012 | 3 | 624397 | 594884 | 59303 | 1278584 |
| 2012 | 4 | 603160 | 578342 | 61696 | 1243198 |
| 2012 | 5 | 645716 | 594905 | 63334 | 1303955 |
| 2012 | 6 | 692473 | 611719 | 78006 | 1382198 |
| 2012 | 7 | 725240 | 599563 | 74460 | 1399263 |
| 2012 | 8 | 753952 | 584090 | 48269 | 1386311 |
| 2012 | 9 | 647908 | 732596 | 71198 | 1451702 |
| 2012 | 10 | 713516 | 675678 | 78502 | 1467696 |
| 2012 | 11 | 746886 | 670960 | 80808 | 1498654 |
| 2012 | 12 | 767251 | 677136 | 88678 | 1533065 |
| 2013 | 1 | 727597 | 684689 | 72037 | 1484323 |
| 2013 | 2 | 639425 | 578097 | 62939 | 1280461 |
| 2013 | 3 | 721778 | 656356 | 69936 | 1448070 |
| 2013 | 4 | 682371 | 620966 | 70297 | 1373634 |
| 2013 | 5 | 746192 | 654916 | 74609 | 1475717 |
| 2013 | 6 | 816634 | 701066 | 75272 | 1592972 |
| 2013 | 7 | 722498 | 567274 | 56494 | 1346266 |
| 2013 | 8 | 801505 | 751406 | 52236 | 1605147 |
| 2013 | 9 | 675544 | 724891 | 66572 | 1467007 |
| 2013 | 10 | 782153 | 711888 | 68806 | 1562847 |
| 2013 | 11 | 733306 | 668666 | 77507 | 1479479 |
| 2013 | 12 | 778264 | 716948 | 72311 | 1567523 |
| 2014 | 1 | 750717 | 693012 | 57888 | 1501617 |
| 2014 | 2 | 590242 | 552074 | 44842 | 1187158 |
| **Tahun** | **Bulan** | **Kedatangan** | **Keberangkatan** | **Transit** | **Total** |
| 2014 | 3 | 671628 | 617490 | 53800 | 1342918 |
| 2014 | 4 | 640629 | 579311 | 52987 | 1272927 |
| 2014 | 5 | 721862 | 625460 | 64573 | 1411895 |
| 2014 | 6 | 781611 | 680664 | 65098 | 1527373 |
| 2014 | 7 | 724498 | 511108 | 45957 | 1281563 |
| 2014 | 8 | 767842 | 823941 | 71186 | 1662969 |
| 2014 | 9 | 692698 | 702141 | 75117 | 1469956 |
| 2014 | 10 | 770043 | 706807 | 73980 | 1550830 |
| 2014 | 11 | 722515 | 658016 | 76505 | 1457036 |
| 2014 | 12 | 817116 | 727690 | 74037 | 1618843 |
| 2015 | 1 | 661875 | 621333 | 65769 | 1348977 |
| 2015 | 2 | 579621 | 516808 | 59084 | 1155513 |
| 2015 | 3 | 623743 | 560909 | 63213 | 1247865 |
| 2015 | 4 | 651219 | 576428 | 61970 | 1289617 |
| 2015 | 5 | 749597 | 630662 | 76223 | 1456482 |
| 2015 | 6 | 735406 | 579568 | 76793 | 1391767 |
| 2015 | 7 | 849308 | 672246 | 64639 | 1586193 |
| 2015 | 8 | 706583 | 793642 | 87004 | 1587229 |
| 2015 | 9 | 685312 | 627371 | 78303 | 1390986 |
| 2015 | 10 | 746425 | 676661 | 87951 | 1511037 |
| 2015 | 11 | 724976 | 651147 | 83084 | 1459207 |
| 2015 | 12 | 868954 | 750387 | 99698 | 1719039 |

*[Halaman ini sengaja dikosongkan]*

# BIODATA PENULIS

Rizka Annisa Kurnia Sari, lahir di Surabaya pada tanggal 24 Juli 1997. Lulus dari SMAN 5 Surabaya pada tahun 2015 dan melanjutkan studi di Departemen Informatika Institut Teknologi Sepuluh Nopember Surabaya. Berpengalaman menjadi asisten dosen mata kuliah matematika informatika, sistem basis data, manajemen basis data, dan analisis perancangan sistem informasi.

Aktif mengikuti organisasi antara lain staf Departemen Media dan Informasi Himpunan Mahasiswa Teknik Computer-Informatik (HMTC) 2016/2017, wakil koordinator 3 bidang web dan kesekretariatan pada Schematics 2016, koordinator bidang web dan kesekretariatan pada Schematics 2017 dan administrator Laboratorium Pemrograman Departemen Informatika ITS.

Dalam menyelesaikan pendidikan sarjana, penulis mengambil tugas akhir bidang minat Dasar dan Terapan Komputasi (DTK) dan juga memiliki ketertarikan di bidang Rancangan Perangkat Lunak (RPL). Penulis dapat dihubungi melalui alamat *e-mail*: pelkenuk.24@gmail.com.